DOCUMENT RESUME

ED 173 443 TM 009 621

AUTHOR Wood, Carolyn J.; Pohland, Paul A.
TITLE Teacher Evaluation: Organizational Maintenance Versus Stimulation of Improved Teaching Performance. The New Mexico Principalship Study.
PUB DATE Apr 79
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Content Analysis; Educational Trends; Effective Teaching; Elementary Secondary Education; *Evaluation Criteria; *Instructional Improvement; *Noninstructional Responsibility; *Rating Scales; Student Teacher Relationship; Teacher Characteristics; *Teacher Evaluation; Teacher Improvement; Teacher Responsibility; *Teacher Role; Teaching Techniques
IDENTIFIERS New Mexico

ABSTRACT

Teacher evaluation procedures appear to focus on organizational maintenance aspects more heavily than on helping teachers improve their teaching performance. This conclusion was reached after a content analysis of teacher evaluation instruments used in New Mexico schools. Items focusing on the instructional role constituted only 28% of the items in the rating scale instruments. Other factors used in teacher evaluations included personal characteristics (30%); administrator/manager role (14%); social role (12%); professional role (7%); organizational membership role (7%); and student achievement (1%). The results also suggested that organizational conservatism and stability, rather than change, appear to be highly valued. From a supplementary analysis of teacher evaluation instruments dating from the turn of the centruy to the present, it was concluded that the secondary emphasis on instructional role has been a persistent characteristic of teacher evaluation over the years. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED173443

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE NEW MEXICO PRINCIPALSHIP STUDY

Teacher Evaluation: Organizational Maintenance Versus Stimulation of Improved Teaching Performance

Paper prepared for the
American Educational Research Association
Annual Meeting
San Francisco, California
April, 1979
Session 10.30

Carolyn J. Wood & Paul A. Pohland

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Carolyn J. Wood

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Department of Educational Administration
College of Education
The University of New Mexico
Albuquerque, New Mexico 87131

TM009 621

> And all the time -- such is the tragi-comedy of our situation-- we continue to clamor for those very qualities we are rendering impossible. You can hardly open a periodical without coming across the statement that what our civilization needs is more 'drive,' or dynamism, or self-sacrifice, or 'creativity.' In a sort of ghastly simplicity we remove the organ and demand the function. . . . We castrate and bid the geldings be fruitful.
>
> -- C. S. Lewis (1947, p. 35)

Recently the Educational Research Service (1978) published the results of a survey conducted to ascertain the current practices of evaluating teacher performance in school systems throughout the United States. One section of the report described "the uses made of the summative evaluation reports" (p.179). Based on the responses from 363 school districts, ERS listed the four most frequently mentioned purposes of teacher evaluation as: 1) to help teachers improve their teaching performance (349 responses); 2) to decide on renewed appointment of probationary teachers (328 responses); 3) to recommend probationary teachers for tenure or continuing contract status (326 responses); and 4) to recommend dismissal of unsatisfactory tenured or continuing contract teachers (317 responses). We suggest that the most frequently listed purpose, i.e., helping teachers improve their teaching performance, is fundamentally different from the other three and, more importantly, that this purpose is not clearly operationalized in the evaluation processes currently used in many school districts. Our stance is derived from the reality of the teacher evaluation processes which focus much more heavily on organizational maintenance than they do on improving teacher classroom

performance. Our justification for this stance is derived from an examination of teacher evaluation instruments, both in terms of the content items on which teachers are assessed and on the procedural and structural properties of the instruments. Following a report of these findings, we will discuss one factor which helps to explain why teacher evaluation processes tend to emphasize organizational maintenance functions rather than improved teaching performance.

Nature of Instruments Used in the Evaluation Process

Content Characteristics of the Rating Scale Instruments

The major research task was an analysis of the teacher evaluation instruments currently used in New Mexico schools. Instruments were solicited from all 88 districts. Of the 71 districts responding, 65 submitted rating scales and 6 submitted open ended instruments. Data from the former only will be reported in this section of the study.

An early decision was to engage in a content analysis (Holsti, 1968) of the instruments. Consequently we first scanned a number of the more complex instruments in order to provisionally identify a set of categories and decision rules. One outcome of this process was the observation that the items in the rating scale instruments focused more heavily on the teachers than upon teaching. Reflected in the items was what appeared to be a major concern for the teacher as a person, a professional, an administrator/manager, and a member of the school organization. Hence we devised a tentative category system based upon roles and role performance.[1] Second, the adequacy of this category system and attendant decision rules

was tested by content analyzing a random sample of ten additional instruments. In the main the role conceptualizations were upheld although necessary refinements became apparent. Third, we re-analyzed all of the data on the basis of the final set of categories and decision rules which emerged in stage two. The results of the analysis of the 65 rating scale instruments are displayed in Figure 1.

- - - - - - - - - - - - - - -

Insert Figure 1 about here

- - - - - - - - - - - - - - -

Given the manifest purpose of teacher evaluation, one might have predicted that the instructional role category would account for the majority of items on the evaluation forms. However, the items in this category (e.g., "adequate preparation of lesson plans"; "has work areas arranged for maximum pupil stimulation and accomplishments"; "challenges students to think, inquire and analyze"; "used varied techniques in evaluation of student progress") accounted for only 28.22% of the items in the rating scale instruments. Personal characteristics of the teacher (e.g., "enthusiastic", "sense of humor", "considerate", "punctual") actually accounted for a higher percentage of items (30.03%). However, as Table 1 suggests, all of the rating scale instruments contained at least one item related to the instructional role category.[2]

- - - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - - -

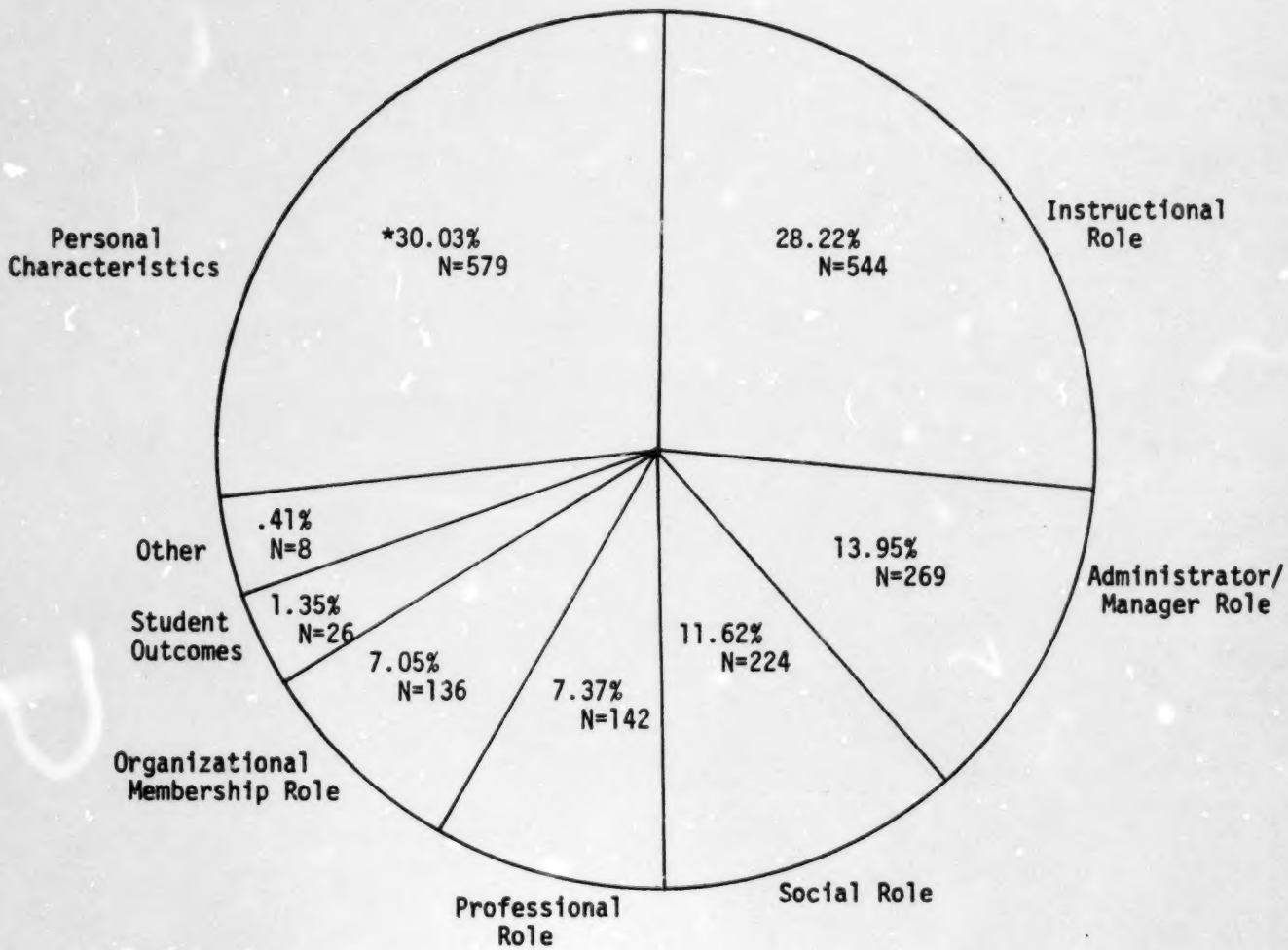

Figure 1: Distribution by category of items in rating scale instruments

*Percentages based on N = 1928 items

Table 1

Frequencies and Percentages of Items by Category in the Rating Scale Instruments

Evaluation Instrument Categories	Number of Districts Reporting	% of Districts Reporting*	Number of Items	% of Items**
Instructional Roles	65	100.00	544	28.22
Personal Characteristics	62	95.38	579	30.03
Social Roles	61	93.85	224	11.62
Administrator/Manager Roles	59	90.77	269	13.95
Organizational Membership Roles	55	84.62	136	7.05
Professional Roles	54	83.08	142	7.37
Student Outcomes	15	23.08	26	1.35
Other	7	10.77	8	.41
Total			1,928	100.00

*Percentages based on N = 65 districts

**Percentages based on N = 1,928 items

In addition to focusing on the teacher rather than on the teaching, the evaluation instruments required the evaluator to take into account and pass judgment on the teacher's entire work world. Items pertaining to the teacher's administrator/manager role, social role, professional role, and organizational membership role collectively accounted for almost 40% of the total number of items on the rating scale instruments. Though the teacher's performance of these roles may be thought necessary to the smooth operation of the school organization, these roles are essentially supportive to the primary role of teaching. A brief description of the types of items included under each of these four role categories follows.

First, almost 14% of the items focused on what we chose to label the teacher's administrator/manager role. Items placed in this category evaluated teachers on their ability: 1) to control students, e.g., "is able to maintain discipline"; 2) to attend to the physical, housekeeping and aesthetic qualities of the classroom environment, e.g., "keeps room neat, attractive, and in order"; and 3) to perform administrative duties, e.g., "keeps records accurately" and "calls for substitute on time".

Second,the social role teachers were expected to perform in the school organization was the focus of almost 12% of the items in the rating scale instruments. By social role we refer to the teacher's interpersonal relationships both within and without the school site. Sixty-one school districts (93.85%) included one or more items which referred to the teacher's relationship with: 1) the community, e.g., "membership in community groups"; 2) other school staff, e.g., "treats non-certified

staff with respect and understanding"; 3) students, e.g., "teacher-pupil relationship"; and 4) district personnel, e.g., "works cooperatively with other school or district personnel".

Third, 54 districts (83.08%) included in their teacher evaluation instruments at least one item which assessed some aspect of the teacher's professional role. By professional role we refer to those items associated with: 1) the teacher's academic preparation and mastery of the subject matter content, e.g., "knows subject matter and books being used"; 3) the teacher's commitment to professional growth and development, e.g., "interest in self-improvement"; and 3) items which judged the ethical standards of the teacher, e.g., "adheres to accepted ethical and social standards of the community and the profession". The items in the professional role category constituted 7.37% of the items in the rating scale instruments.

Finally, the items categorized under the teacher's organizational membership role (N = 136; 7.05%) indicated that teachers were expected to exhibit attitudes and behaviors in concert with the needs and goals of the school organization. Items under this rubric assessed the degree to which the teacher 1) was "law-abiding", e.g., "seeks to understand and abide by all school rules and regulations"; 2) complied with the suggestions of other professionals, e.g., "accepts and uses constructive criticism"; 3) assumed non-teaching responsibilities, e.g., "assumes assignments and extra duties willingly"; and 4) exhibited and organizational perspective, e.g., "sees his/her responsibility in relation to the total school program".

In sum, school districts appeared to place a high value on what we conceptualized as organizational maintenance roles. Almost 40% (39.99%) of the items addressed such issues as the teacher's ability to maintain order and function efficiently, to get along well with others, to exhibit "professional" behaviors, and to comply with the rules of the organization. Our major point is that while such roles and role behaviors may well be essential for organizational maintenance, they are only tangentially related to classroom teaching.

The organizational maintenance emphasis was further accented in one of the personal characteristics sub-categories. Through a more fine-grained analysis of the sub-category "personal traits,"[3] four groupings emerged: 1) traits related to membership in the organization, 2) traits related to work with students, 3) traits related to the task of teaching, and 4) traits related to the teacher as a person. Table 2 displays the 398 items in the evaluation instruments referring to these groupings.

- - - - - - - - - - - - - - - -

Insert Table 2 about here

- - - - - - - - - - - - - - - -

The data from Table 2 indicate that those traits associated with organizational membership (Grouping 1) combined with those related to the teacher as a person (Grouping 4) account for 60.06% of the personal traits assessed by the school districts. It is of more than casual interest to note further that those traits more closely associated with

Table 2

Frequency of the Inclusion of Specific Personal Traits Related to Organizational Membership, Students, the Task of Teaching and the Teacher as a Person

Personal Traits	Number of Times Reported	Personal Traits	Number of Times Reported
1. Traits Related to Organizational Membership		3. Traits Related to Task of Teaching	
Punctual/prompt	68	Enthusiastic	43
Cooperative	12	Initiative	15
Dependable	12	Resourceful	6
Accurate	11	Leadership	4
Responsible	10	Creative	2
Loyal	4	Original	2
Adaptible	4	Innovative	1
Flexible	2	Dedicated	1
Total	123	Total	74
2. Traits Related to Work with Students		4. Traits Related to the Teacher as a Person	
Interested in	22	Judgment	29
Fair	17	Sense of Humor	16
Patient	11	Self-controlled	11
Compassionate	10	Tactful	10
Positive Attitude Toward	7	Poised	10
		Sincere	7
Impartial	6	Emotionally Stable	5
Considerate	3	Honest	4
Tolerant	2	Exemplar	3
Sympathetic	2	Calm	3
Constructive	1	Self-confident	3
Democratic	1	Mature	2
Empathetic	1	Friendly	2
Humane	1	Discreet	2
Courteous	1	Other[a]	9
Total	85	Total	116

[a]Each of the following personal traits was reported by one school district: Alert, Perseverant, Patriotic, Open-Minded, Reasonable, Optimistic, Faces Problems Honestly, Good Mental Health, Positive Self-Concept.

the teacher's instructional role, i.e., work with students (Grouping 2) and the task of teaching (Grouping 3) constitute only 39.95% of the total. More particularly, for example, references to punctuality (N = 68) exceed by a factor of three references to interest in students (N = 22).

The inferences from data like these are clear. First, it appears that teacher evaluation instruments reflect a primary concern for ogranizational maintenance and a secondary concern for improvement of instruction. The items in the instruments concentrate heavily on outside of the classroom phenomena. Second, and related to the first, organizational conservatism and stability rather than change appear most highly valued. Traits such as punctuality, cooperation, dependability, responsiblity, accuracy, loyalty and adaptability (N = 121) reflect the former, while traits such as creativity, innovativeness, originality, leadership, resourcefulness, initiative and flexibility (N = 32) mirror the latter and are assessed much less frequently.

In drawing these inferences we are <u>not</u> suggesting that more items should focus on the teacher's in-class instructional role and fewer on what we have identified as the broader organizational membership role. Indeed, the data gathered on current teacher evaluation instruments may be highly pertinent to the maintenance functions of school organizations particularly with reference to making personnel decisions. However we contend equally strongly that maintenance emphasizing data currently obtained under the broad rubric of teacher evaluation are largely irrelevant to the manifest purpose of improving instruction and hence provide little leverage for affecting change in teaching performance.

Supplementary Analysis

We combined the open ended instruments with the rating scales (N = 71) to accomplish an analysis of the design characateristics of the evaluation instruments, i.e., their procedural requirements and structural properties. In this supplementary analysis, we found supporting evidence for the primacy of the organizational maintenance function of teacher evaluation and the corresponding subordination of the improvement of instruction function. First, less than a third of all the documents (23 of 71 districts; 32.39%) indicated specific conferencing requirements. Of these, three (4.23%) required a pre-conference; 19 (26.76%) required a post-conference; and one (1.41%) required both. The more typical pattern was the simple requirement that both the teacher and the evaluator sign the evaluation document. Even here, however, some discrepancy was noted: evaluator signatures were specified requirements on 66 of the 71 documents (92.96%), while somewhat fewer teacher signatures were so required (64 of 71; 90.14%). Furthermore, it might be noted that only one district required conferencing with tenured teachers following a classroom observation.

Second, with reference to the number of observations or classroom visits on which the evaluation was to be based, data from only 11 of the 71 districts (15.49%) suggested that more than one classroom visit was required. Third, it appeared that non-tenured teachers were the primary target of observation.[4] Finally, of the 58 rating scale instruments which used evaluative designations such as "needs to improve," "satisfactory," and so forth, only 24 (41.38%) included designations more positive than "satisfactory," "acceptable," or "average."

Our major point is that if indeed the purpose of teacher evaluation is to improve instruction then it does not seem unreasonable that conferences, both pre- and post-observation, should be routinely scheduled with all teachers observed; that multiple classroom observations should be required of the evaluator; that all teachers, tenured and non-tenured alike, should be involved regularly in the process; and that scaling should be precise, broad range, and indicative of a concern for excellence. These design characteristics were in scant evidence. Their omission communicates to teachers and evaluators alike more of a concern for maintenance and meeting "minimum standards" than systematically improving instruction.

Factors Which Perpetuate the Organizational Maintenance Function of Teacher Evaluation

How does one account for the "tragi-comedy" of the situation where school districts profess that the primary purpose of evaluation is to improve teaching performance whereas the tools and procedures through which they attempt to accomplish this purpose tend to communicate a manifestly different concern? A useful perspective is provided by Platt (1973) and Culbert (1974) who conceptualize troublesome problems in terms of "trap" comments. Platt defines "social traps" in the following manner:

> The term refers to situations in society that contain traps formally like a fish trap, where men or organizations or whole societies get themselves started in some direction or some set of relationships that later prove to be unpleasant or lethal and that they see no easy way to back out of or to avoid. (p. 641)

Culbert's notion of the "organization trap" suggests that traps take the form of assumptions people make about how the organization operates and about how they operate within the organization. "We assume that what others

imply is happening is in fact happening" (p. 12). Furthermore, Culbert indicates that these assumptions "limit the extent to which we manage our organizational lives and subject us to excessive influence by the system" (p.20).

We posit that similar "trap" components exist in teacher evaluation. We will approach our discussion of these traps from the perspective of what Lortie (1975) has labelled "the hand of history" by demonstrating that the evaluation instruments used in the late 1970's reflect emphases similar to those emanating from the age of scientific management. This is initially consistent with Davis' (1964) conclusion of fifteen years ago that "Methods of evaluating teacher competence currently used in public school systems have evolved from practices of many years ago" (p.41).[5]

Empirical support for this stance was obtained by examining item content, procedural characteristics, scaling patterns, and the stated purposes of teacher evaluation instruments from the turn of the century to the present. More specifically, a comparison was made of four instruments drawn from three points in time: the Age of Efficiency (two instruments), the Age of Human Relations (one instrument) and the present time (one instrument). Additionally, comparative analyses were done on summary studies completed during those three time periods. The rationale for instrument selection follows.

The first instrument, Elliott's "Provisional Plan for the Measure of Merit of Teachers," was originally developed in 1910. A later edition, "not materially different from the first" (Boyce, 1915, p. 78), appeared as an Appendix to the 14th Yearbook of the National Society for

the Study of Education in 1915. Boyce further indicates that Elliott's "tentative scheme for measuring teaching efficiency" was "widely used as a pattern" (p. 78).

Boyce's (1915) own instrument is our second exemplar of the Age of Efficiency. Appearing originally in the March 1915 issues of the American School Board Journal, it received further endorsement in the 1915 NSSE Yearbook. Callahan (1962) observes that it was "warmly received . . . and adopted in many school systems" (p. 105), while Davis (1964) notes that it was "so widely copied that it now appears commonplace" (p.47).

The third instrument was utilized by the Oakland Public Schools in the 1930's. Reavis and Cooper (1945) identify it as one of the "examples of the better rating practices" (p.46) since as a probationary teacher instrument it could be instrumental in "protecting the school from the [permanent] appointment of unpromising teachers" (p. 57). The final instrument chosen for comparative analysis is one of the more popular (N = 11) currently in use in New Mexico schools.

In making a comparative analysis it became necessary to utilize a common set of categories and decision rules.[6] Consequently, we re-analyzed the item content of the Elliott, Boyce, and Oakland instruments using the New Mexico format. The results are displayed in Table 3.

- - - - - - - - - - - - - - - -

Insert Table 3 about here

- - - - - - - - - - - - - - - -

Table 3

Frequency and Percentages of Items by Category
in Four Rating Scale Instrument Used
During Three Comparison Periods

Categories derived from the Wood & Pohland Study	Age of Efficiency: Elliott (1914) Frequency	Elliott (1914) Percent	Boyce (1915) Frequency	Boyce (1915) Percent	Age of Human Relations: Reavis & Cooper (1945) Frequency	Reavis & Cooper (1945) Percent	Current Practice: Wood & Pohland (1978) Frequency	Wood & Pohland (1978) Percent
1. Personal Characteristics	19	34.55	18	36.00	14	46.67	14	22.58
2. Instructional Role	10	18.18	13	26.00	5	16.67	17	27.42
3. Admin/Mgr/Role	5	9.09	4	8.00	1	3.33	16	25.81
4. Social Role	6	10.91	4	8.00	3	10.00	6	9.68
5. Professional Role	6	10.91	5	10.00	4	13.33	4	6.45
6. Organizational Membership Role	6	10.91	3	6.00	2	6.67	5	8.06
7. Student Outcome	3	5.45	3	6.00	1	3.33	0	0.00
TOTAL	55	100.00	50	100.00	30	100.00	62	100.00

Additionally, Boyce, Reavis and Cooper, and we had analyzed sets of instruments (Boyce, 50 rating schemes, 25 categories, 395 items; Reavis and Cooper, 85 rating scales, 7 categories, 1539 items; Wood and Pohland, 65 rating scales, 8 categories, 1928 items). Therefore, we did a comparison of the summarized findings. Again, we applied the New Mexico category system and decision rules to the earlier data. The results of this comparative analysis are presented in Table 4.

- - - - - - - - - - - - - - -

Insert Table 4 about here

- - - - - - - - - - - - - - -

The "hand of history" is clearly noticeable in the data presented in Tables 3 and 4. First, the importance first ascribed by Elliott to the assessment of personal characteristics has persisted and has in fact risen overall (Table 4). Only in the single instrument analyzed by Wood and Pohland is the percent of items assessing personal characteristics less than any other category (Table 3). As expected, the primacy of assessing personnel characteristics reached its nadir in the Age of Human Relations (Table 3). Its after effects are probably what accounts for the high percentage figure today (Table 4).

That the importance of assessing this factor has endured over the years may well be attributed to the ascribed socialization function of schools (Bidwell, 1965), the associated demand for exemplary adult role models (Lortie, 1975), and the historic dominance of an Anglo-Saxon value orientation. Hence the persistence of such "personal characteristics" as self-control, promptness, reliability, initiative, and the like.

Table 4

Frequency and Percentage of Items by Category in Rating Scale Instruments Used During Three Comparison Periods

	Age of Efficiency Boyce (1915)		Age of Human Relations Reavis & Cooper (1945)		Current Practice Wood & Pohland (1978)	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Personal Characteristics	99	25.06	396	25.73	579	30.03
Instructional Role	80	20.25	309	20.08	544	28.22
Administrator Manager Role	74	18.73	174	11.31	269	13.95
Social Role	19	4.81	226	14.68	224	11.62
Professional Role	67	16.96	186	12.09	142	7.37
Organizational Membership Role	48	12.15	88	5.72	136	7.05
Student Outcomes	8	2.03	146	9.49	26	1.35
Other	0	0.00	14	.91	8	.41
TOTAL	395	99.99	1539	100.01	1928	100.00

Second, the relative emphasis given to the instructional role of the teacher (Table 4) has remained fairly consistent, showing but a slight percentage decrease in the Age of Human Relations and a slight percentage increase today. Even in Table 3 where the widest fluctuations occur, the variations within era (Elliott vs. Boyce) are not appreciably different than the variations between eras (Reavis and Cooper vs. Wood and Pohland). Further, the stability of this factor is reflected in the nature of the items assessed. For example, Boyce's (1915) "Efficiency Record" rated teachers on "Grasp of subject-matter;" the Oakland instrument reported by Reavis and Cooper (1945) assessed "Knowledge of Subject Matter;" and the New Mexico instrument most commonly used rates teachers on "Knows subject matter and books being used."

The data are somewhat less clear relative to the assessment of teacher administrative/managerial competence. Table 3 indicates a moderate emphasis on those aspects of the teacher's role in the 1910's, a decreased emphasis as expected in the 40's, and a major percentage increase in the 70's. The latter phenomenon is totally reasonable given the current press for accountability. Table 4, however, shows a marked emphasis in that domain during the Age of Efficiency, a phenomenon consistent with Callahan's (1962) analysis; an equally marked diminution in the Age of Human Relations; and a slow trend for re-emphasizing that element of teacher behavior today.

Somewhat similar ambiguities surround the assessment of the teacher's social role. One might have predicted that the instruments used during the Age of Human Relations would tend to have exaggerated the teacher's social role. Table 4 indicates that during the Age of Human Relations

there was a rather pronounced rise in the number of items related to the social role when compared with the number for the Age of Efficiency and furtner suggests a moderate emphasis in present day instruments. That this emphasis is not reflected in Table 3 is somewhat puzzling. In Table 3 the comparison of individual rating scale instruments during the three time periods yields virtually no differences in the percentage of items related to the teacher's social role. Here as elsewhere, however, the explanation may well reside in the choice of instrument. In that sense Table 4 data may present a more accurate picture of overall trends.

One of the more interesting comparisons is relative to the data on the assessment of the professional role. When one looks at the end points of Tables 3 and 4 the trend is clear -- less attention to that role. This may well support Culbert's (1974) point that school people make an assumption about what is happening within the organization and act as if it were indeed happening. The assumption in this case is that the increased professionalization/training of teachers (Blood, 1978) results in increased proficiency and consequently one need pay less attention to the professional role issue. There is, however, a mid-point anomaly which should be pointed out. Table 3 indicates an increase in attention to this variable during the Age of Human Relations when compared with the Age of Efficiency while Table 4 shows the reverse. Several observations can be made. One, the variation between the two sets of Age of Human Relations data is slight (13.33 vs. 12.09). Second, the difference may again be artifactual (one instrument vs. 85); third, and more intriguingly, it may reflect the tensions in the role of principal as evaluator which arose during the 1940's. On the one hand

principals were urged to re-define themselves as instructional leaders while on the other hand they were exhorted to maintain their administrative orientation. Our data from this study are insufficient to resolve this issue here.

The final two categories, Organizational Membership Role and Student Outcomes, provide clear cut trend data. The organizational membership role of the teacher received moderate assessment emphasis during the Age of Efficiency, weak emphasis in the Age of Human Relations, and is currently undergoing a resurgence, although not to the degree of prominence it enjoyed during the Age of Efficiency (Tables 3 and 4). Student outcome data has never been seriously used as a measure of teaching competence. It was a minor factor in the Age of Efficiency, rose moderately during the Age of Human Relations, and all but disappears in the sample 1970's. instrument (Table 3). At the least, one can conclude from this that thrusts for accountability and/or "competency based teaching" have not been echold in evaluation practices.

In sum, relative to the content of teacher evaluation instruments, the "hand of history" appears very heavy-handed indeed. Variables upon which teachers are evaluated appear to be largely constant over time with but minor variations and fluctuations in emphases. In the main, the conception of teacher competence established at the turn of the century remains intact.

Much the same observation can be made relative to certain procedural characteristics. For example, when we asked, "Who does the evaluating?" the historical pattern persisted. Boyce (1915) reports that "rating the

teachers is done in one way or another by the principals or superintendents or both" (p. 15). Reavis and Cooper (1945) note that 89.3% of the raters were principals, supervisors, or "administrative and supervisory staff" (p. 39). Davis (1964) states bluntly that "the principal is the primary rating officer" (p. 37). New Mexico data confirms that condition. State statutes require principals to engage in teacher evaluations for the purpose of making personnel recommendations to the superintendent. In brief, administrative practice begun in the early 20th century have been legitimated by law.

The method of rating teachers has also changed but little. Rating scales were prevalent in Boyce's day and clearly advocated by him. Reavis and Cooper reported that in one study of 104 school districts all but one used some form of rating system (p. 19). In our New Mexico study, 65 of 71 districts (91.55%) used some form of rating scale.

It is also interesting to note that scaling patterns have remained fairly stable. As Table 5 indicates, two, three, or four point scales

- - - - - - - - - - - - - -

Insert Table 5 about here

- - - - - - - - - - - - - -

accounted for 58.68% of all scales in 1915 and 73.85% of all scales in 1978. What the Table shows is a clear present trend toward reducing the number of scale units. Five point scales prevalent in the mid-forties are now relatively rare, and two and three point scales alone account for 46.16% of all scales currently in use in New Mexico. The resolution of the problem originally posed by Boyce, i.e., the difficulty in discriminating between scale units, appears to have been resolved by simply reducing the number of scale values.

Table 5

Interval Scale Values Used During

the Three Comparison Periods

Number of Interval Scale Values	Age of Efficiency Boyce (1915)			Age of Human Relations Reavis & Cooper (1945)			Current Practice Wood & Pohland (1978)		
	Frequency	Percent		Frequency	Percent		Frequency	Percent	
2	4	3.31		10	11.76		14	21.54	
3	20	16.53	58.68%	12	14.12	42.35%	16	24.62	73.85%
4	47	38.84		14	16.47		18	27.69	
5	20	16.53		36	42.35		10	15.38	
6-20	9	7.44		9	10.59		5	7.69	
Variable	0	0.00		3	3.53		0	0.00	
Indeterminant	0	0.00		1	1.18		0	0.00	
Other	21	17.36		0	0.00		2	3.08	
TOTAL	121	100.01		85	100.00		65	100.00	

Finally, we looked at the stated purposes for engaging in teacher evaluation. Here again we have come full circle. Professor Elliott (in Boyce, 1915) in a "Special Note" which followed the presentation of his "Provisional Plan for the Measure of Merit of Teachers" stated unambiguously:

> It is believed that this analytical plan for the study of teaching merit will be of the greatest service if placed directly in the hands of teachers for their own guidance, and as a basis for cooperative effort between teachers and supervisors. The plan is not intended to be used as a score card by inspectors. (p. 81; emphasis in original)

Elliott's intentions were quickly subverted. Boyce reports the four major uses of teachers ratings as "(1) for the private information of the superintendent; (2) for the purpose of furnishing a basis for salary adjustment, or to control salary adjustment; (3) for use in connection with other factors, such as length of service, examination, or special work to determine promotion; (4) to improve the teachers in service" (p. 23). Reavis and Cooper's (1945) analysis of 85 ratings scales yielded 16 purposes, the top four being (1) for re-employment (N = 23; 27.1%); (2) to change teaching assignment (N = 14; 16.5%); (3) to transfer to a different school (N = 11; 12.9%); and (4) to determine salary (N = 6;7.1%). All other categories of purpose, including instructional imporvement, had frequencies of less than five. The major point is that the teacher self-improvement advocated by Elliott had literally disappeared by the mid forties. Curiously, however, that thrust re-appears in the seventies. In our New Mexico study, the four most commonly listed purposes were: 1) to improve the quality of instruction (N = 11); 2) to assist the teacher

in identifying areas for improvement (N = 6); 3) to improve employee performance (N = 5); and 4) to measure or evaluate teacher performance/effectiveness/competence (N = 5).[7] Additionally, in the opening paragraph of this report we cited the results of the 1978 Educational Research Services Study which posited a similar emphasis on improving teaching performance.

Again, Culbert's "organization traps" may be at work. Having become propagandized to the point of believing principals are instructional leaders and that the purpose of teacher evaluation is the improvement of instruction we assume those conditions to be true. Our analyses suggest rather forcefully that they are not.

Summary

Initially we proposed that the manifest purpose of teacher evaluation as articulated by school districts was to improve instruction but that the latent purpose was to enhance organizational maintenance. From our comparative analsis of teacher evaluation instruments from three points in time corresponding roughly to the Age of Efficiency, the Age of Human Relations, and the 1970's we are prepared to talk about the persistence of the rhetoric and the reality. The rhetoric of evaluation for improvement was originally articulated by Elliott, although perhaps not self-consciously so. It persists. The reality of teacher evaluation as a mechanism for personnel decision making -- a key element in organizational maintenance -- was established by Boyce. It too persists, as strongly or more strongly than ever. Together the rhetoric and reality constitute the hand of history, the trap, from which few school districts have been able to escape.

Footnotes

[1]Two categories, "Student Outcomes" and "Other", do not fit into the conceptual scheme of teachers' roles or role performance. However, these categories account for only 1.76% of the content items.

[2]The variance in the number of items devoted to the instructional role on the rating scale instruments may be of interest to the reader. The number of items on which a teacher was rated ranged from 6 to 63. Two instruments included one item and one instrument included 27 items related to the teacher's instructional role. The modal number of items was five (assessed in ten instruments) and the median was seven. The "average" instrument contained eight items related to the teacher's instructional role.

[3]The sub-category personal traits contained 398 items and thus represented 68.74% of the personal characteristics category and 20.64% of the total number of items in the rating scale instruments. Items included under five other sub-categories accounted for the remaining 31.26% of the personal characteristics category: 1) voice/speech (12.61%), 2) grooming (8.98%), 3) physical fitness (6.39%), 4) general characteristics/attitudes (2.76%) and 5) the residual category "other" (.52%).

[4]Our data on this point was limited since 50 of the 71 instruments made no indication of the target of evaluation (tenured vs non-tenured teachers). However, supporting documentation provided by some school districts, e.g., policy manuals, indicated a clear distinction in evaluation requirements for those two groups.

[5]Subsequently we would argue that such "evolution" (if there has been evolution at all!) is on a Darwinian scale.

[6]An example of a decision rule was to code double headed items, e.g., "optimism and enthusiasm," as separate items. This of necessity altered the total N. For instance, the Elliott instrument contained 44 items; however, by coding double headed items as single items, the number of items increased to 55.

[7]From one to three purposes of teacher evaluation were listed by each of the nineteen school districts.

References

Bidwell, C. E. The school as a formal organization. In J. G. March (Ed.), Handbook of organizations. Chicago: Rand McNally, 1965.

Blood, R. E. Factors affecting the principalship yesterday and today. In The principalship: 1978. Santa Fe, New Mexico: Legislative School Study Committee, 1978.

Boyce, A. C. Methods of measuring teachers' efficiency. 14th Yearbook, National Society for the Study of Education, Part 2. Chicago: University of Chicago Press, 1915.

Callahan, R. E. Education and the cult of efficiency. Chicago: University of Chicago Press, 1962.

Culbert, S. A. The organization trap. New York: Basic Books, 1974.

Davis, H. Evolution of current practices in evaluating teacher competence. In B. J. Biddle & W. J. Ellena (Eds.), Contemporary research on teacher effectiveness. New York: Holt, Rinehart and Winston, 1964.

Educational Research Service. Evaluating teacher performance. Arlington, Virginia: Educational Research Service, 1978.

Holsti, O. R. Content analysis. In G. Lindzey & E. Aronson (Eds.), The handbook of social psychology. Vol. II (2d ed.) Reading, Mass.: Addison-Wesley, 1968.

Lewis, C. S. The abolition of man. New York: Macmillan, 1947.

Lortie, C. D. Schoolteacher: A sociological study. Chicago: University of Chicago Press, 1975.

Platt, J. Social traps. American Psychologist, 1973, 28 641-651.

Reavis, W. C. & Cooper, D. H. Evaluation of teacher merit in city school systems. Supplementary Educational Monographs, Number 59. Chicago: University of Chicago, 1945.